Project Idea

The goal of this project is to predict house prices using data collected from web scraping and machine learning techniques.

1. **Data Collection:**
   We collect data about house prices by scraping websites. The scraped data includes the most important features, such as:
   - Price
   - Taxes
   - Number of rooms
   - Address
   - Garage availability
   - Size
2. **Data Preprocessing and Cleaning:**
   After collecting the data, we preprocess and clean it to ensure its accuracy and usability for analysis.
3. **Statistical Analysis:**
   - We perform regression analysis to identify key patterns and relationships in the data.
   - Statistical methods like the Partial F-Test are used to determine the most significant features that influence house prices.
   - Lasso and Ridge regularization techniques are applied to refine the model and address multicollinearity issues.
4. **Model Building:**
   Using the processed data, we build a machine learning model to predict house prices.
5. **Database Integration:**
   Predicted house prices are stored in a database (MongoDB) for future reference and analysis.
6. **Clustering Analysis:**
   We use unsupervised learning methods, such as K-Means clustering, to group houses based on their price levels.
7. **Visualization:**
   Finally, we create visualizations using Power BI to present insights and trends in the data.